



---

# Journal of Statistical Software

August 2014, Volume 59, Issue 6.

<http://www.jstatsoft.org/>

---

## phtt: Panel Data Analysis with Heterogeneous Time Trends in R

Oualid Bada  
University of Bonn

Dominik Liebl  
Université Libre de Bruxelles

---

### Abstract

The R package **phtt** provides estimation procedures for panel data with large dimensions  $n$ ,  $T$ , and general forms of unobservable heterogeneous effects. Particularly, the estimation procedures are those of Bai (2009) and Kneip, Sickles, and Song (2012), which complement one another very well: both models assume the unobservable heterogeneous effects to have a factor structure. Kneip *et al.* (2012) considers the case in which the time-varying common factors have relatively smooth patterns including strongly positively auto-correlated stationary as well as non-stationary factors, whereas the method of Bai (2009) focuses on stochastic bounded factors such as ARMA processes. Additionally, the **phtt** package provides a wide range of dimensionality criteria in order to estimate the number of the unobserved factors simultaneously with the remaining model parameters.

*Keywords:* panel data, unobserved heterogeneity, principal component analysis, factor dimension, interactive fixed effects.

---

## 1. Introduction

One of the main difficulties and at the same time appealing advantages of panel models is their need to deal with the problem of the unobserved heterogeneity. Classical panel models, such as *fixed effects* or *random effects* models, try to model unobserved heterogeneity using dummy variables or structural assumptions on the error term (see, e.g., Baltagi 2005). In both cases the unobserved heterogeneity is assumed to remain constant over time within each cross-sectional unit – apart from an eventual common time trend. This assumption might be reasonable for approximating panel data with fairly small temporal dimensions  $T$ ; however, for panel data with large  $T$  this assumption becomes very often implausible.

Nowadays, the availability of panel data with large cross-sectional dimensions  $n$  and large time dimensions  $T$  has triggered the development of a new class of panel data models. Recent

discussions by [Ahn, Lee, and Schmidt \(2013\)](#), [Pesaran \(2006\)](#), [Bai \(2009\)](#), [Bai, Kao, and Ng \(2009\)](#), and [Kneip \*et al.\* \(2012\)](#) have focused on advanced panel models for which the unobservable individual effects are allowed to have heterogeneous (i.e., individual specific) time trends that can be approximated by a factor structure. The basic form of this new class of panel models can be presented as follows:

$$y_{it} = \sum_{j=1}^P x_{itj} \beta_j + \nu_{it} + \epsilon_{it} \quad \text{for } i \in \{1, \dots, n\} \text{ and } t \in \{1, \dots, T\}, \quad (1)$$

where  $y_{it}$  is the dependent variable for each individual  $i$  at time  $t$ ,  $x_{itj}$  is the  $j$ th element of the vector of explanatory variables  $x_{it} \in \mathbb{R}^P$ , and  $\epsilon_{it}$  is the idiosyncratic error term. The time-varying individual effects  $\nu_{it} \in \mathbb{R}$  of individual  $i$  for the time points  $t \in \{1, \dots, T\}$  are assumed to be generated by  $d$  common time-varying factors. The following two specifications of the time-varying individual effects  $\nu_{it}$  are implemented in our R package **phtt**:

$$\nu_{it} = \begin{cases} v_{it} &= \sum_{l=1}^d \lambda_{il} f_{lt}, & \text{for the model of } \text{Bai (2009)}, \\ v_i(t) &= \sum_{l=1}^d \lambda_{il} f_l(t), & \text{for the model of } \text{Kneip *et al.* (2012)}. \end{cases} \quad (2)$$

Here,  $\lambda_{il}$  are unobserved individual loadings parameters,  $f_{lt}$  are unobserved common factors for the model of [Bai \(2009\)](#),  $f_l(t)$  are the unobserved common factors for the model of [Kneip \*et al.\* \(2012\)](#), and  $d$  is the unknown factor dimension.

Note that the explicit consideration of an intercept in model (1) is not necessary but may facilitate interpretation. If  $x_{it}$  includes an intercept, the time-varying individual effects  $\nu_{it}$  are centered around zero. If  $x_{it}$  does not include an intercept, the time-varying individual effects  $\nu_{it}$  are centered around the overall mean.

Model (1) includes the classical panel data models with additive time-invariant individual effects and common time-specific effects. This model is obtained by choosing  $d = 2$  with a first common factor  $f_{1t} = 1$  for all  $t \in \{1, \dots, T\}$  that has individual loadings parameters  $\lambda_{i1}$ , and a second common factor  $f_{2t}$  that has the same loadings parameter  $\lambda_{i2} = 1$  for all  $i \in \{1, \dots, n\}$ .

An intrinsic problem of factor models lies in the fact that the true factors are only identifiable up to rotation. In order to ensure the uniqueness of these parameters, a number of  $d^2$  restrictions are required. The usual normalization conditions are given by

$$\begin{aligned} \text{(a)} \quad & \frac{1}{T} \sum_{t=1}^T f_{lt}^2 = 1 \quad \text{for all } l \in \{1, \dots, d\}, \\ \text{(b)} \quad & \sum_{t=1}^T f_{lt} f_{kt} = 0 \quad \text{for all } l, k \in \{1, \dots, d\} \text{ with } k \neq l, \text{ and} \\ \text{(c)} \quad & \sum_{i=1}^n \lambda_{il} \lambda_{ik} = 0 \quad \text{for all } l, k \in \{1, \dots, d\} \text{ with } k \neq l; \end{aligned}$$

see, e.g., [Bai \(2009\)](#) and [Kneip \*et al.\* \(2012\)](#). For the model of [Kneip \*et al.\* \(2012\)](#),  $f_{lt}$  in conditions (a) and (b) has to be replaced by  $f_l(t)$ . As usual in factor models, a certain degree of indeterminacy remains, because the factors can only be determined up to sign changes and different ordering schemes.

[Kneip \*et al.\* \(2012\)](#) consider the case in which the common factors  $f_l(t)$  show relatively smooth patterns over time. This includes strongly positively auto-correlated stationary as well as non-stationary factors. The authors propose to approximate the time-varying individual

effects  $v_i(t)$  by smooth nonparametric functions, say,  $\vartheta_i(t)$ . In this way (1) becomes a semi-parametric model and its estimation is done using a two-step estimation procedure, which we explain in more detail in Section 2. The asymptotic properties of this method rely, however, on independent and identically distributed errors.

Alternatively, Bai (2009) allows for weak forms of heteroskedasticity and dependency in both time and cross-section dimensions and proposes an iterated least squares approach to estimate (1) for stationary time-varying individual effects  $v_{it}$  such as ARMA processes or non-stationary deterministic trends. However, Bai (2009) rules out a large class of non-stationary processes such as stochastic processes with integration.

Moreover, Bai (2009) assumes the factor dimension  $d$  to be a known parameter, which is usually not the case. Therefore, the **phtt** package uses an algorithmic refinement of Bai's method proposed by Bada and Kneip (2014) in order to estimate the number of unobserved common factors  $d$  jointly with the remaining model parameters; see Section 4 for more details. Besides the implementations of the methods proposed by Kneip *et al.* (2012), Bai (2009), and Bada and Kneip (2014) the R package **phtt** comes with a wide range of criteria (16 in total) for estimating the factor dimension  $d$ . The main functions of the **phtt** package are given in the following list:

- **KSS()**: Computes the estimators of the model parameters according to the method of Kneip *et al.* (2012); see Section 2.
- **Eup()**: Computes the estimators of the model parameters according to the method of Bai (2009) and Bada and Kneip (2014); see Section 4.
- **OptDim()**: Allows for a comparison of the estimated factor dimensions  $\hat{d}$  obtained from many different (in total 16) criteria; see Section 3.
- **checkSpecif()**: Tests whether to use a classical fixed effects panel model or a panel model with individual effects  $\nu_{it}$ ; see Section 5.1.

For the returned objects the usual **print()**, **summary()**, **plot()**, **coef()** and **residuals()** methods are provided.

Standard methods for estimating models for panel and longitudinal data are also implemented in the R (R Core Team 2014) packages **plm** (Croissant and Millo 2008), **nlme** (Pinheiro, Bates, DebRoy, Sarkar, and R Core Team 2014), and **lme4** (Bates, Maechler, Bolker, and Walker 2014); see Croissant and Millo (2008) for an exhaustive comparison of these packages. Recently, Millo and Piras (2012) published the R package **splm** for spatial panel data models. The **phtt** package further extends the toolbox for statisticians and econometricians and provides the possibility of analyzing panel data with large dimensions  $n$  and  $T$  and can be considered when the unobserved heterogeneity effects are time-varying.

To the best of our knowledge, our **phtt** package (Bada and Liebl 2014) is the first software package that offers the estimation methods of Bai (2009) and Kneip *et al.* (2012). Regarding the different dimensionality criteria that can be accessed via the function **OptDim()** only those of Bai and Ng (2002) are publicly available as MATLAB codes (The MathWorks Inc. 2012) from the homepage of Serena Ng (<http://www.columbia.edu/~sn2294/>).

To demonstrate the use of our functions, we re-explore the well known **Cigar** dataset, which is frequently used in the literature of panel models. The panel contains the per capita cigarette

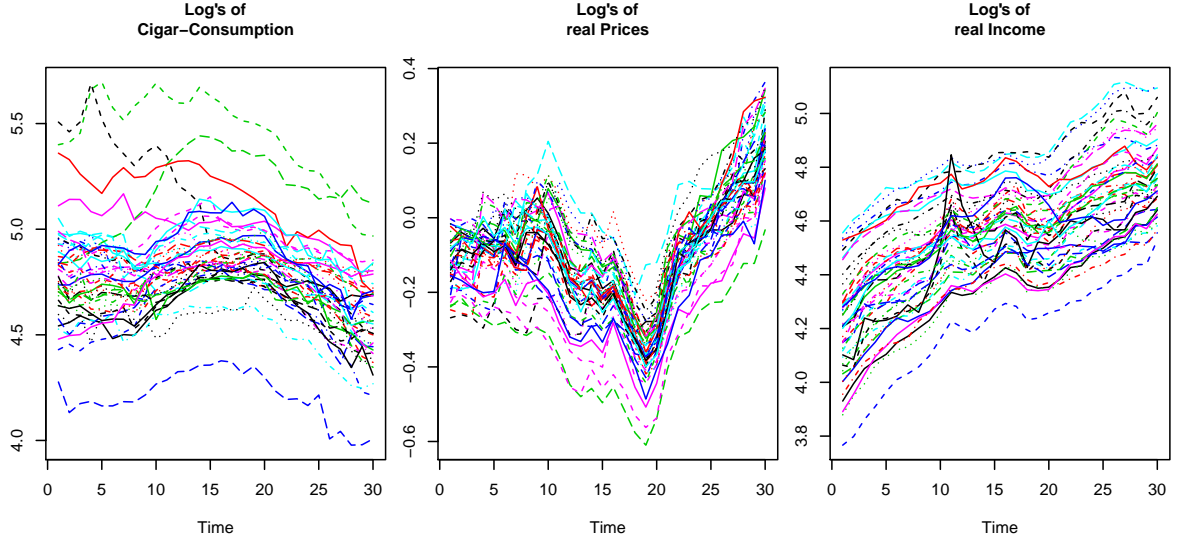


Figure 1: Plots of the dependent variable  $\ln(\text{Consumption}_{it})$  and regressor variables  $\ln(\text{Price}_{it})$  and  $\ln(\text{Income}_{it})$ .

consumptions of  $n = 46$  American states from 1963 to 1992 ( $T = 30$ ) as well as data about the income per capita and cigarette prices (see, e.g., Baltagi and Levin 1986, for more details on the dataset).

We follow Baltagi and Li (2004), who estimate the following panel model:

$$\ln(\text{Consumption}_{it}) = \mu + \beta_1 \ln(\text{Price}_{it}) + \beta_2 \ln(\text{Income}_{it}) + e_{it}. \quad (3)$$

Here,  $\text{Consumption}_{it}$  presents the sales of cigarettes (packs of cigarettes per capita),  $\text{Price}_{it}$  is the average real retail price of cigarettes, and  $\text{Income}_{it}$  is the real disposable income per capita. The index  $i \in \{1, \dots, 46\}$  denotes the single states and the index  $t \in \{1, \dots, 30\}$  denotes the year.

We revisit this model, but allow for a multidimensional factor structure such that

$$e_{it} = \sum_{l=1}^d \lambda_{il} f_{lt} + \epsilon_{it}.$$

The `Cigar` dataset can be obtained from the `phtt` package using the function `data("Cigar", package = "phtt")`. In Figure 1 the panels of the variables  $\ln(\text{Consumption}_{it})$ ,  $\ln(\text{Price}_{it})$ , and  $\ln(\text{Income}_{it})$  are shown.

Section 2 is devoted to a short introduction of the method of Kneip *et al.* (2012), which is appropriate for relatively smooth common factors  $f_l(t)$ . Section 3 presents the usage of the function `OptDim()`, which provides access to a wide range of panel dimensionality criteria recently discussed in the literature on factor models. Section 4 deals with the explanation as well as application of the panel method proposed by Bai (2009), which is basically appropriate for stationary and relatively unstructured common factors  $f_{lt}$ .

## 2. Panel models for heterogeneity in time trends

The panel model proposed by [Kneip et al. \(2012\)](#) can be presented as follows:

$$y_{it} = \sum_{j=1}^P x_{itj} \beta_j + v_i(t) + \epsilon_{it}, \quad (4)$$

where the time-varying individual effects  $v_i(t)$  are parametrized in terms of common non-parametric basis functions  $f_1(t), \dots, f_d(t)$  such that

$$v_i(t) = \sum_{l=1}^d \lambda_{il} f_l(t). \quad (5)$$

The asymptotic properties of this method rely on second order differences of  $v_i(t)$ , which apply to continuous functions as well as to classical discrete stochastic time series processes such as (S)AR(I)MA processes. Therefore, the functional notation of the time-varying individual effects  $v_i(t)$  and their underlying common factors  $f_1(t), \dots, f_d(t)$  does not restrict them to a purely functional interpretation. The main idea of this approach is to approximate the time series of individual effects  $v_i(t)$  by smooth functions  $\vartheta_i(t)$ .

The estimation approach proposed by [Kneip et al. \(2012\)](#) relies on a two-step procedure: first, estimates of the common slope parameters  $\beta_j$  and the time-varying individual effects  $v_i(t)$  are obtained semi-parametrically. Second, functional principal component analysis is used to estimate the common factors  $f_1(t), \dots, f_d(t)$ , and to re-estimate the time-varying individual effects  $v_i(t)$  more efficiently. In the following we describe both steps in more detail.

**Step 1:** The unobserved parameters  $\beta_j$  and  $v_i(t)$  are estimated by the minimization of

$$\sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \left( y_{it} - \sum_{j=1}^P x_{itj} \beta_j - \vartheta_i(t) \right)^2 + \sum_{i=1}^n \kappa \int_1^T \frac{1}{T} \left( \vartheta_i^{(m)}(s) \right)^2 ds, \quad (6)$$

over all  $\beta_j \in \mathbb{R}$  and all  $m$ -times continuously differentiable functions  $\vartheta_i(t)$ , where  $\vartheta_i^{(m)}(t)$  denotes the  $m$ th derivative of the function  $\vartheta_i(t)$ . A first approximation of  $v_i(t)$  is then given by  $\tilde{v}_i(t) := \hat{\vartheta}_i(t)$ . Spline theory implies that any solution  $\hat{\vartheta}_i(t)$  possesses an expansion in terms of a natural spline basis  $z_1(t), \dots, z_T(t)$  such that  $\hat{\vartheta}_i(t) = \sum_{s=1}^T \hat{\zeta}_{is} z_s(t)$ ; see, e.g., [Boor \(2001\)](#). Using the latter expression, we can rewrite (6) to formalize the following objective function:

$$S(\beta, \zeta) = \sum_{i=1}^n \left( \|Y_i - X_i \beta - Z \zeta_i\|^2 + \kappa \zeta_i^\top R \zeta_i \right), \quad (7)$$

where  $Y_i = (y_{i1}, \dots, y_{iT})^\top$ ,  $X_i = (x_{i1}^\top, \dots, x_{iT}^\top)^\top$ ,  $\beta = (\beta_1, \dots, \beta_P)^\top$ ,  $\zeta_i = (\zeta_{i1}, \dots, \zeta_{iT})^\top$ .  $Z$  and  $R$  are  $T \times T$  matrices where  $Z$  has elements  $\{z_s(t)\}_{s,t=1,\dots,T}$  and  $R$  elements  $\{\int z_s^{(m)}(t) z_k^{(m)}(t) dt\}_{s,k=1,\dots,T}$ , respectively.  $\kappa$  is a preselected smoothing parameter to control the smoothness of  $\hat{\vartheta}_i(t)$ . We follow the usual choice of  $m = 2$ , which leads to cubic smoothing splines.

In contrast to [Kneip et al. \(2012\)](#), we do not specify a common time effect in model (4), but the vector of explanatory variables is allowed to contain an intercept. This means that the time-varying individual effects  $v_i(t)$  are not centered around zero for each specific time point  $t$ , but around a common intercept term. The separate estimation of the common time effect, say  $\theta_t$ , is also possible with our **phtt** package; we discuss this in detail in Section 5.

The semi-parametric estimators  $\hat{\beta}, \hat{\zeta}_i = (\hat{\zeta}_{i1}, \dots, \hat{\zeta}_{iT})^\top$ , and  $\tilde{v}_i = (\tilde{v}_{i1}, \dots, \tilde{v}_{iT})^\top$  can be obtained by minimizing  $S(\beta, \zeta)$  over all  $\beta \in \mathbb{R}^P$  and  $\zeta \in \mathbb{R}^{T \times n}$ .

The solutions are given by

$$\hat{\beta} = \left( \sum_{i=1}^n X_i^\top (I - \mathcal{Z}_\kappa) X_i \right)^{-1} \left( \sum_{i=1}^n X_i^\top (I - \mathcal{Z}_\kappa) Y_i \right), \quad (8)$$

$$\hat{\zeta}_i = (Z^\top Z + \kappa R)^{-1} Z^\top (Y_i - X_i \hat{\beta}), \text{ and} \quad (9)$$

$$\tilde{v}_i = \mathcal{Z}_\kappa (Y_i - X_i \hat{\beta}), \text{ where } \mathcal{Z}_\kappa = Z (Z^\top Z + \kappa R)^{-1} Z^\top. \quad (10)$$

**Step 2:** The common factors are obtained by the first  $d$  eigenvectors  $\hat{\gamma}_1, \dots, \hat{\gamma}_d$  that correspond to the largest eigenvalues  $\hat{\rho}_1, \dots, \hat{\rho}_d$  of the empirical covariance matrix

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \tilde{v}_i \tilde{v}_i^\top. \quad (11)$$

The estimator of the common factor  $f_l(t)$  is then defined by the  $l$ th scaled eigenvector

$$\hat{f}_l(t) = \sqrt{T} \hat{\gamma}_{lt} \text{ for all } l \in \{1, \dots, d\}, \quad (12)$$

where  $\hat{\gamma}_{lt}$  is the  $t$ th element of the eigenvector  $\hat{\gamma}_l$ . The scaling factor  $\sqrt{T}$  yields that  $\hat{f}_l(t)$  satisfies the normalization condition  $\frac{1}{T} \sum_{t=1}^T \hat{f}_l(t)^2 = 1$  as listed above in Section 1. The estimates of the individual loadings parameters  $\lambda_{il}$  are obtained by ordinary least squares regressions of  $(Y_i - X_i \hat{\beta})$  on  $\hat{f}_l$ , where  $\hat{f}_l = (\hat{f}_l(1), \dots, \hat{f}_l(T))^\top$ . Recall from conditions (a) and (b) that  $\hat{\lambda}_{il}$  can be calculated as follows:

$$\hat{\lambda}_{il} = \frac{1}{T} \hat{f}_l^\top (Y_i - X_i \hat{\beta}). \quad (13)$$

A crucial part of the estimation procedure of [Kneip et al. \(2012\)](#) is the re-estimation of the time-varying individual effects  $v_i(t)$  in Step 2 by  $\hat{v}_i(t) := \sum_{l=1}^d \hat{\lambda}_{il} \hat{f}_l(t)$ , where the factor dimension  $d$  can be determined, e.g., by the sequential testing procedure of [Kneip et al. \(2012\)](#) or by any other dimensionality criterion; see also Section 3. This re-estimation leads to more efficiently estimated time-varying individual effects.

[Kneip et al. \(2012\)](#) derive the consistency of the estimators as  $n, T \rightarrow \infty$  and show that the asymptotic distribution of common slope estimators is given by  $\hat{\Sigma}_\beta^{-1/2} (\hat{\beta} - E_\epsilon(\hat{\beta})) \xrightarrow{d} \mathbf{N}(0, I)$ , where

$$\hat{\Sigma}_\beta = \sigma^2 \left( \sum_{i=1}^n X_i^\top (I - \mathcal{Z}_\kappa) X_i \right)^{-1} \left( \sum_{i=1}^n X_i^\top (I - \mathcal{Z}_\kappa)^2 X_i \right) \left( \sum_{i=1}^n X_i^\top (I - \mathcal{Z}_\kappa) X_i \right)^{-1}. \quad (14)$$

A consistent estimator of  $\sigma^2$  can be obtained by

$$\hat{\sigma}^2 = \frac{1}{(n-1)T} \sum_{i=1}^n \|Y_i - X_i \hat{\beta} - \sum_{l=1}^{\hat{d}} \hat{\lambda}_{il} \hat{f}_l\|^2. \quad (15)$$

To determine the optimal smoothing parameter  $\kappa_{opt}$ , [Kneip \*et al.\* \(2012\)](#) propose the following cross validation (CV) criterion:

$$CV(\kappa) = \sum_{i=1}^n \|Y_i - X_i \hat{\beta}_{-i} - \sum_{l=1}^d \hat{\lambda}_{-i,l} \hat{f}_{-i,l}\|^2, \quad (16)$$

where  $\hat{\beta}_{-i}$ ,  $\hat{\lambda}_{-i,l}$ , and  $\hat{f}_{-i,l}$  are estimates of the parameters  $\beta$ ,  $\lambda$ , and  $f_l$  based on the dataset without the  $i$ th observation. Unfortunately, this criterion is computationally very costly and requires determining the factor dimension  $d$  in advance. To overcome this disadvantage, we propose a plug-in smoothing parameter that is discussed in more detail in the following [Section 2.1](#).

## 2.1. Computational details

Theoretically, it is possible to determine  $\kappa$  by the CV criterion in (16); however, cross validation is computationally very costly. Moreover, [Kneip \*et al.\* \(2012\)](#) do not explain how the factor dimension  $d$  is to be specified during the optimization process, which is critical since the estimator  $\hat{d}$  is influenced by the choice of  $\kappa$ .

In order to get a quick and effective solution, we propose to determine the smoothing parameter  $\kappa$  by generalized cross validation (GCV). However, we cannot apply the classical GCV formulas as proposed, e.g., in [Craven and Wahba \(1978\)](#) since we do not know the parameters  $\beta$  and  $v_i(t)$ . Our computational algorithm for determining the GCV smoothing parameter  $\kappa_{GCV}$  is based on the method of [Cao and Ramsay \(2010\)](#), who propose optimizing objective functions of the form (7) by updating the parameters iteratively in a functional hierarchy. Formally, the iteration algorithm can be described as follows:

1. For given  $\kappa$  and  $\beta$ , we optimize (7) with respect to  $\zeta_i$  to get

$$\hat{\zeta}_i = (Z^\top Z + \kappa R)^{-1} Z^\top (Y_i - X_i \beta). \quad (17)$$

2. By using (17), we minimize (7) with respect to  $\beta$  to get

$$\hat{\beta} = \left( \sum_{i=1}^n X_i^\top X_i \right)^{-1} \left( \sum_{i=1}^n X_i^\top (Y_i - Z \hat{\zeta}_i) \right) \quad (18)$$

3. Once (17) and (18) are obtained, we optimize the following GCV criterion to calculate  $\kappa_{GCV}$ :

$$\kappa_{GCV} = \arg \min_{\kappa} \frac{1}{\frac{n}{T} \text{tr}(I - \mathcal{Z}_\kappa)^2} \sum_{i=1}^n \|Y_i - X_i \hat{\beta} - \mathcal{Z}_\kappa (Y_i - X_i \hat{\beta})\|^2. \quad (19)$$



The program starts with initial estimates of  $\beta$  and  $\kappa$  and proceeds with Steps 1, 2, and 3 in recurrence until convergence of all parameters. The initial value  $\hat{\beta}_{\text{start}}$  is defined in (49) and the initial value  $\kappa_{\text{start}}$  is the GCV-smoothing parameter of the residuals  $Y_i - X_i \hat{\beta}_{\text{start}}$ .

The advantage of this approach is that the inversion of the  $P \times P$  matrix in (18) does not have to be updated during the iteration process. Moreover, the determination of the GCV minimizer in (19) can be easily performed in R using the function `smooth.spline()`, which calls a rapid C routine.

But note that the GCV smoothing parameter  $\kappa_{GCV}$  in (19) does not explicitly account for the factor structure of the time-varying individual effects  $v_i(t)$  as formalized in (2). In fact, given that the assumption of a factor structure is true, the goal should not be to obtain optimal estimates of  $v_i(t)$  but rather to obtain optimal estimates of the common factors  $f_l(t)$ , which implies that the optimal smoothing parameter  $\kappa_{\text{opt}}$  will be smaller than  $\kappa_{GCV}$ ; see Kneip *et al.* (2012).

If the goal is to obtain optimal estimates of  $f_l(t)$ ,  $\kappa_{\text{opt}}$  will be used as an upper bound when minimizing the CV criterion (16) (via setting the argument `CV = TRUE`); which, however, can take some time. Note that, this optimal smoothing parameter  $\kappa_{\text{opt}}$  depends on the unknown factor dimension  $d$ . Therefore, we propose to, first, estimate the dimension based on the smoothing parameter  $\kappa_{GCV}$  and, second, to use the estimated dimension  $\hat{d}$  (via explicitly setting the dimension argument `factor.dim =  $\hat{d}$` ) in order to determine the dimension-specific smoothing parameter  $\kappa_{\text{opt}}$  (via setting the argument `CV = TRUE`).

## 2.2. Application

This section is devoted to the application of the method of Kneip *et al.* (2012) discussed above. The computation of this method is accessible through the function `KSS()`, which has the following arguments:

```
KSS(formula, additive.effects = c("none", "individual", "time", "twoways"),
    consult.dim.crit = FALSE, d.max = NULL, sig2.hat = NULL, factor.dim = NULL,
    level = 0.01, spar = NULL, CV = FALSE, convergence = 1e-06,
    restrict.mode = c("restrict.factors", "restrict.loadings"), ...)
```

The argument `formula` is compatible with the usual R-specific symbolic specification of the model. The unique specificity here is that the variables should be defined as  $T \times n$  matrices, where  $T$  is the temporal dimension and  $n$  is the number of the cross-section units.<sup>1</sup>

The argument `additive.effects` makes it possible to extend model (4) with additional additive `individual`, `time`, or `twoways` effects as discussed in Section 5.

If the logical argument `consult.dim.crit` is set to `TRUE` all dimensionality criteria discussed in Section 3 are computed and the user is asked to choose one of their results.

The arguments `d.max` and `sig2.hat` are required for the computation of some dimensionality criteria discussed in Section 3. If their default values are maintained, the function internally computes `d.max =  $\lfloor \min\{\sqrt{n}, \sqrt{T}\} \rfloor$`  and `sig2.hat` as in (15), where  $\lfloor x \rfloor$  indicates the integer part of  $x$ . The argument `level` allows to adjust the significance level for the dimensionality testing procedure (21) of Kneip *et al.* (2012); see Section 3.

<sup>1</sup>Note that **phtt** is written for balanced panels. Missing values have to be replaced in a pre-processing step by appropriate imputation methods.



`CV` is a logical argument. If it is set to `TRUE` the cross validation criterion (16) of Kneip *et al.* (2012) will be computed. In the default case, the function uses the GCV method discussed in Section 2.1.

The factor dimension  $d$  can be pre-specified by the argument `factor.dim`. Recall from restriction (a) that  $\frac{1}{T} \sum_{t=1}^T \hat{f}_l(t)^2 = 1$ .

Alternatively, it is possible to standardize the individual loadings parameters such that  $\frac{1}{n} \sum_{i=1}^n \hat{\lambda}_{il} = 1$ , which can be done by setting `restrict.mode = "restrict.loadings"`.

As an illustration we estimate the cigarettes model (3) introduced in Section 1:

$$\ln(\text{Consumption}_{it}) = \mu + \beta_1 \ln(\text{Price}_{it}) + \beta_2 \ln(\text{Income}_{it}) + e_{it} \quad (20)$$

with

$$e_{it} = \sum_{l=1}^d \lambda_{il} f_l(t) + \epsilon_{it},$$

In the following lines of code we load the `Cigar` dataset and take logarithms of the three variables, `Consumptionit`, `Priceit/cpit` and `Incomeit/cpit`, where `cpit` is the consumer price index. The variables are stored as  $T \times n$  matrices. This is necessary, because the `formula` argument of the `KSS()` function takes the panel variables as matrices in which the number of rows has to be equal to the temporal dimension  $T$  and the number of columns has to be equal to the individual dimension  $n$ .

```
R> library("phtt")
R> data("Cigar", package = "phtt")
R> n <- 46; T <- 30
R> l.Consumption <- log(matrix(Cigar$sales, T, n))
R> cpi <- matrix(Cigar$cpi, T, n)
R> l.Price <- log(matrix(Cigar$price, T, n)/cpi)
R> l.Income <- log(matrix(Cigar$ndi, T, n)/cpi)
```

The model parameters  $\beta_1$ ,  $\beta_2$ , the factors  $f_l(t)$ , the loadings parameters  $\lambda_{il}$ , and the factor dimension  $d$  can be estimated by the `KSS()` function with its default arguments. Inferences about the slope parameters can be obtained using the `summary()` method.

```
R> Cigar.KSS <- KSS(formula = l.Consumption ~ l.Price + l.Income)
R> (Cigar.KSS.summary <- summary(Cigar.KSS))
```

Call:

```
KSS.default(formula = l.Consumption ~ l.Price + l.Income)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.11	-0.01	0.00	0.01	0.12

Slope-Coefficients:

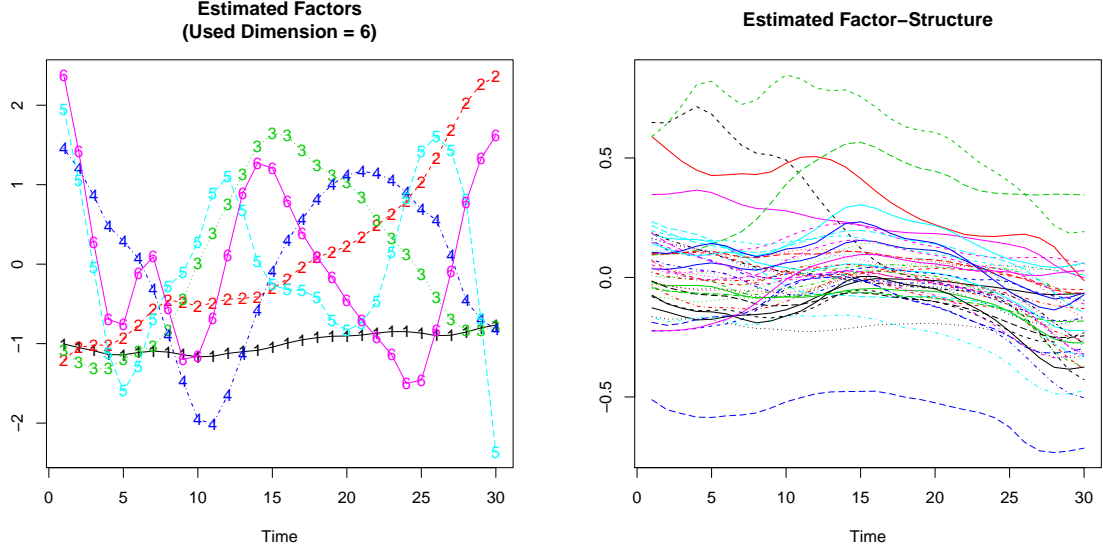


Figure 2: Left panel: Estimated factors  $\hat{f}_1(t), \dots, \hat{f}_6(t)$ . Right panel: Estimated time-varying individual effects  $\hat{v}_1(t), \dots, \hat{v}_n(t)$ .

	Estimate	StdErr	z.value	Pr(>z)
(Intercept)	4.0600	0.1770	23.00	< 2.2e-16 ***
l.Price	-0.2600	0.0223	-11.70	< 2.2e-16 ***
l.Income	0.1550	0.0382	4.05	5.17e-05 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Additive Effects Type: none

Used Dimension of the Unobserved Factors: 6

Residual standard error: 0.000725 on 921 degrees of freedom

R-squared: 0.99

The effects of the log-real prices for cigarettes and the log-real incomes on the log-sales of cigarettes are highly significant and in line with results in the literature. The summary output reports an estimated factor dimension of  $\hat{d} = 6$ . In order to get a visual impression of the six estimated common factors  $\hat{f}_1(t), \dots, \hat{f}_6(t)$  and the estimated time-varying individual effects  $\hat{v}_1(t), \dots, \hat{v}_n(t)$ , we provide a `plot()` method for the objects of class `'summary.KSS'` which are returned by the `summary()` method of the `'KSS'` objects returned by `KSS()`.

```
R> plot(Cigar.KSS.summary)
```

The left panel of Figure 2 shows the six estimated common factors  $\hat{f}_1(t), \dots, \hat{f}_6(t)$  and the right panel of Figure 2 shows the  $n = 46$  estimated time-varying individual effects  $\hat{v}_1(t), \dots, \hat{v}_n(t)$ . The common factors are ordered according to the decreasing sequence of their eigenval-

ues. Obviously, the first common factor is nearly time-invariant; this suggests extending the model (20) by additive `individual` (time-invariant) effects; see Section 5 for more details.

By setting the logical argument `consult.dim.crit = TRUE`, the user can choose from other dimensionality criteria, which are discussed in Section 3. Note that the consideration of different factor dimensions  $d$  would not alter the results for the slope parameters  $\beta$  since the estimation procedure of Kneip *et al.* (2012) for the slope parameters  $\beta$  does not depend on the dimensionality parameter  $d$ .

### 3. Panel criteria for selecting the number of factors

In order to estimate the factor dimension  $d$ , Kneip *et al.* (2012) propose a sequential testing procedure based on the following test statistic:

$$\text{KSS}(d) = \frac{n \sum_{r=d+1}^T \hat{\rho}_r - (n-1) \hat{\sigma}^2 \text{tr}(\mathcal{Z}_\kappa \hat{\mathcal{P}}_d \mathcal{Z}_\kappa)}{\hat{\sigma}^2 \sqrt{2n \cdot \text{tr}((\mathcal{Z}_\kappa \hat{\mathcal{P}}_d \mathcal{Z}_\kappa)^2)}} \stackrel{a}{\sim} N(0, 1), \quad (21)$$

where  $\hat{\mathcal{P}}_d = I - \frac{1}{T} \sum_{l=1}^d f_l f_l^\top$  with  $f_l = (f_l(1), \dots, f_l(T))^\top$ , and

$$\hat{\sigma}^2 = \frac{1}{(n-1) \text{tr}((I - \mathcal{Z}_\kappa)^2)} \sum_{i=1}^n \|(I - \mathcal{Z}_\kappa)(Y_i - X_i \hat{\beta})\|^2. \quad (22)$$

The selection method can be described as follows: choose a significance level  $\alpha$  (e.g.,  $\alpha = 1\%$ ) and begin with  $H_0 : d = 0$ . Test if  $\text{KSS}(0) \leq z_{1-\alpha}$ , where  $z_{1-\alpha}$  is the  $(1 - \alpha)$ -quantile of the standard normal distribution. If the null hypothesis can be rejected, go on with  $d = 1, 2, 3, \dots$  until  $H_0$  cannot be rejected. Finally, the estimated dimension is then given by the smallest dimension  $d$ , which leads to a rejection of  $H_0$ .

The dimensionality criterion of Kneip *et al.* (2012) can be used for stationary as well as non-stationary factors. However, this selection procedure has a tendency to ignore factors that are weakly auto-correlated. As a result, the number of factors can be underestimated.

More robust against this kind of underestimation are the criteria of Bai and Ng (2002). The basic idea of their approach consists simply of finding a suitable penalty term  $g_{nT}$ , which countersteers the undesired variance reduction caused by an increasing number of factors  $\hat{d}$ . Formally,  $\hat{d}$  can be obtained by minimizing the following criterion:

$$\text{PC}(l) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{y}_{it}(l))^2 + l g_{nT} \quad (23)$$

for all  $l \in \{1, 2, \dots\}$ , where  $\hat{y}_{it}(l)$  is the fitted value for a given factor dimension  $l$ . To estimate consistently the dimension of stationary factors Bai and Ng (2002) propose specifying  $g_{nT}$  by

one of the following penalty terms:

$$g_{nT}^{(\text{PC1})} = \hat{\sigma}^2 \frac{(n+T)}{nT} \log \left( \frac{nT}{n+T} \right), \quad (24)$$

$$g_{nT}^{(\text{PC2})} = \hat{\sigma}^2 \frac{(n+T)}{nT} \log(\min\{n, T\}), \quad (25)$$

$$g_{nT}^{(\text{PC3})} = \hat{\sigma}^2 \frac{\log(\min\{n, T\})}{\min\{n, T\}}, \text{ and} \quad (26)$$

$$g_{nT}^{(\text{BIC3})} = \hat{\sigma}^2 \frac{(n+T-l)}{nT} \log(nT), \quad (27)$$

where  $\hat{\sigma}^2$  is the sample variance estimator of the residuals  $\hat{\epsilon}_{it}$ . The proposed criteria are denoted by PC1, PC2, PC3, and BIC3 respectively. Note that only the first three criteria satisfy the requirements of Theorem 2 in [Bai and Ng \(2002\)](#), i.e., (i)  $g_{nT} \rightarrow 0$  and (ii)  $\min\{n, T\}g_{nt} \rightarrow \infty$ , as  $n, T \rightarrow \infty$ . These conditions ensure consistency of the selection procedure without imposing additional restrictions on the proportional behavior of  $n$  and  $T$ . The requirement (i) is not always fulfilled for BIC3, especially when  $n$  is too large relative to  $T$  or  $T$  is too large relative to  $n$  (e.g.,  $n = \exp(T)$  or  $T = \exp(n)$ ). In practice, BIC3 seems to perform very well, especially when the idiosyncratic errors are cross-correlated.

The variance estimator  $\hat{\sigma}^2$  can be obtained by

$$\hat{\sigma}^2(d_{\max}) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{y}_{it}(d_{\max}))^2, \quad (28)$$

where  $d_{\max}$  is an arbitrary maximal dimension that is larger than  $d$ . This kind of variance estimation can, however, be inappropriate in some cases, especially when  $\hat{\sigma}^2(d_{\max})$  underestimates the true variance. To overcome this problem, [Bai and Ng \(2002\)](#) propose three additional criteria (IC1, IC2, and IC3):

$$\text{IC}(l) = \log \left( \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{y}_{it}(l))^2 \right) + l g_{nT} \quad (29)$$

with

$$g_{nT}^{(\text{IC1})} = \frac{(n+T)}{nT} \log \left( \frac{nT}{n+T} \right), \quad (30)$$

$$g_{nT}^{(\text{IC2})} = \frac{(n+T)}{nT} \log(\min\{n, T\}), \text{ and} \quad (31)$$

$$g_{nT}^{(\text{IC3})} = \frac{\log(\min\{n, T\})}{\min\{n, T\}}. \quad (32)$$

In order to improve the finite sample performance of IC1 and IC2, [Alessi, Barigozzi, and Capasso \(2010\)](#) propose to multiply the penalties  $g_{nT}^{(\text{IC1})}$  and  $g_{nT}^{(\text{IC2})}$  with a positive constant  $c$  and apply the calibration strategy of [Hallin and Liška \(2007\)](#). The choice of  $c$  is based on the inspection of the criterion behavior through  $J$ -different tuples of  $n$  and  $T$ , i.e.,  $(n_1, T_1), \dots, (n_J, T_J)$ , and for different values of  $c$  in a pre-specified grid interval. We denote the refined criteria in our package by ABC.IC1 and ABC.IC2 respectively. Note that such a modification does not affect the asymptotic properties of the dimensionality estimator.

Under similar assumptions, [Ahn and Horenstein \(2013\)](#) propose selecting  $d$  by maximizing the ratio of adjacent eigenvalues (or the ratio of their growth rate). The criteria are referred to as *eigenvalue ratio* (ER) and *growth ratio* (GR) and defined as following:

$$\text{ER} = \frac{\hat{\rho}_l}{\hat{\rho}_{l+1}}, \quad (33)$$

$$\text{GR} = \frac{\log \left( \sum_{r=l}^T \hat{\rho}_r / \sum_{r=l+1}^T \hat{\rho}_r \right)}{\log \left( \sum_{r=l+1}^T \hat{\rho}_r / \sum_{r=l+2}^T \hat{\rho}_r \right)}. \quad (34)$$

Note that the theory of the above dimensionality criteria PC1, PC2, PC3, BIC3, IC1, IC2, IC3, IPC1, IPC2, IPC3, ABC.IC1, ABC.IC2, KSS.C, ER, and GR are developed for stochastically bounded factors. In order to estimate the number of unit root factors, [Bai \(2004\)](#) proposes the following panel criteria:

$$\text{IPC}(l) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{y}_{it}(l))^2 + l g_{nT}, \quad (35)$$

where

$$g_{nT}^{(\text{IPC1})} = \hat{\sigma}^2 \frac{\log(\log(T))}{T} \frac{(n+T)}{nT} \log \left( \frac{nT}{n+T} \right), \quad (36)$$

$$g_{nT}^{(\text{IPC2})} = \hat{\sigma}^2 \frac{\log(\log(T))}{T} \frac{(n+T)}{nT} \log(\min\{n, T\}), \text{ and} \quad (37)$$

$$g_{nT}^{(\text{IPC3})} = \hat{\sigma}^2 \frac{\log(\log(T))}{T} \frac{(n+T-l)}{nT} \log(nT). \quad (38)$$

Alternatively, [Onatski \(2010\)](#) introduced a threshold approach based on the empirical distribution of the sample covariance eigenvalues, which can be used for both stationary and non-stationary factors. The estimated dimension is obtained by

$$\hat{d} = \max\{l \leq d_{\max} : \hat{\rho}_l - \hat{\rho}_{l-1} \geq \delta\},$$

where  $\delta$  is a positive threshold, estimated iteratively from the data. We refer to this criterion as ED, which stands for eigenvalue differences.

### 3.1. Application

The dimensionality criteria introduced above are implemented in the function `OptDim()`, which has the following arguments:

```
OptDim(Obj, criteria = c("PC1", "PC2", "PC3", "BIC3", "IC1", "IC2", "IC3",
  "IPC1", "IPC2", "IPC3", "ABC.IC1", "ABC.IC2", "KSS.C", "ED", "ER", "GR"),
  standardize = FALSE, d.max, sig2.hat, spar, level = 0.01,
  c.grid = seq(0, 5, length.out = 128), T.seq, n.seq)
```

The desired criteria can be selected by one or several of the following character variables: "KSS.C", "PC1", "PC2", "PC3", "BIC2", "IC1", "IC2", "IC3", "ABC.IC1", "ABC.IC2", "ER",

"GR", "IPC1", "IPC2", "IPC3", and "ED". The default significance level used for the "KSS" criterion is `level = 0.01`. The values of  $d_{\max}$  and  $\hat{\sigma}^2$  can be specified externally by the arguments `d.max` and `sig2.hat`. By default, `d.max` is computed internally as  $d_{\max} = \lfloor \min\{\sqrt{n}, \sqrt{T}\} \rfloor$  and `sig2.hat` as in (22) and (28). The arguments `c.grid`, `T.seq`, and `n.seq` are required for computing "ABC.IC1" and "ABC.IC2". The grid interval of the calibration parameter can be externally specified with `c.grid`. The  $J$ -tuples,  $(n_1, T_1), \dots, (n_J, T_J)$ , can be specified by using appropriate vectors in `T.seq`, and `n.seq`. If these two arguments are left unspecified, the function constructs internally the following sequences:  $T - C, T - C + 1, \dots, T$ , and  $n - C, n - C + 1, \dots, n$ , for  $C = \min\{\sqrt{n}, \sqrt{T}, 30\}$ . Alternatively, the user can specify only the length of the sequences by giving appropriate integers to the arguments `T.seq`, and `n.seq`, to control for  $C$ .

The input variable can be standardized by choosing `standardize = TRUE`. In this case, the calculation of the eigenvalues is based on the correlation matrix instead of the covariance matrix for all criteria.

As an illustration, imagine that we are interested in the estimation of the factor dimension of the variable  $\ln(\text{Consumption}_{it})$  with the dimensionality criterion "PC1". The function `OptDim()` requires a  $T \times n$  matrix as input variable.

```
R> OptDim(Obj = l.Consumption, criteria = "PC1")
```

```
Call: OptDim.default(Obj = l.Consumption, criteria = "PC1")
```

```
-----
```

```
Criterion of Bai and Ng (2002):
```

```
PC1
  5
```

`OptDim()` offers the possibility of comparing the results of different selection procedures by giving the corresponding criteria to the argument `criteria`. If the argument `criteria` is left unspecified, `OptDim()` automatically compares all 16 procedures.

```
R> (OptDim.obj <- OptDim(Obj = l.Consumption, criteria = c("PC3", "ER",
+ "GR", "IPC1", "IPC2", "IPC3"), standardize = TRUE))
```

```
Call: OptDim.default(Obj = l.Consumption, criteria = c("PC3", "ER",
+ "GR", "IPC1", "IPC2", "IPC3"), standardize = TRUE)
```

```
-----
```

```
Criterion of Bai and Ng (2002):
```

```
PC3
  5
```

```
-----
```

```
Criteria of Ahn and Horenstein (2013):
```

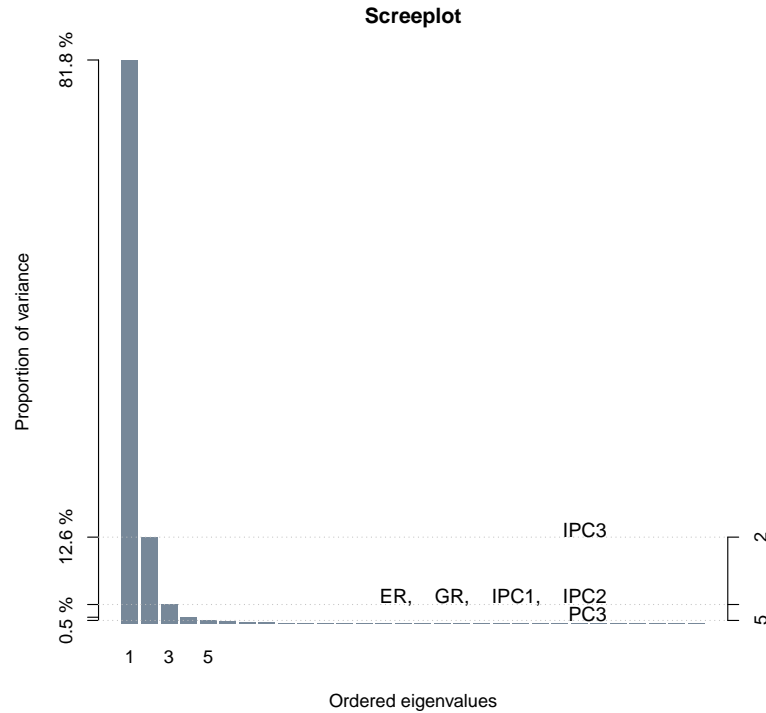


Figure 3: Scree plot produced by the `plot()` method for ‘`OptDim`’ objects. Most of the dimensionality criteria (ER, GR, IPC1 and IPC2) suggest using the dimension  $\hat{d} = 3$ .

```

ER GR
 3  3

-----
Criteria of Bai (2004):

IPC1 IPC2 IPC3
  3    3    2

```

In order to help users to choose the most appropriate dimensionality criterion for the data, ‘`OptDim`’ objects are provided with a `plot()` method. This method displays, in descending order, the magnitude of the eigenvalues in percentage of the total variance and indicates where the selected criteria detect the dimension; see Figure 3.

```
R> plot(OptDim.obj)
```

We, now, come back to the `KSS()` function, which offers an additional way to compare the results of all dimensionality criteria and to select one of them: If the `KSS()` argument `consult.dim = TRUE`, the results of the dimensionality criteria are printed on the console of R and the user is asked to choose one of the results.



```
R> KSS(formula = l.Consumption ~ -1 + l.Price + l.Income, consult.dim = TRUE)
```

```
-----
Results of Dimension-Estimations
```

```
-Bai and Ng (2002):
```

PC1	PC2	PC3	BIC3	IC1	IC2	IC3
5	5	5	4	5	5	5

```
-Bai (2004):
```

IPC1	IPC2	IPC3
3	3	2

```
-Alessi et al. (2010):
```

ABC.IC1	ABC.IC2
3	3

```
-Kneip et al. (2012):
```

KSS.C
6

```
-Onatski (2009):
```

ED
3

```
-Ahn and Horenstein (2013):
```

ER	GR
3	6

```
-----
Please, choose one of the proposed integers:
```

After entering a number of factors, e.g., 6 we get the following feedback:

```
Used dimension of unobs. factor structure is: 6
-----
```

Note that the maximum number of factors that can be given, cannot exceed the highest estimated factor dimension (here the maximal dimension would be 6). A higher dimension can be chosen using the argument `factor.dim`.

## 4. Panel models with stochastically bounded factors

The panel model proposed by [Bai \(2009\)](#) can be presented as follows:

$$y_{it} = \sum_{j=1}^P x_{itj} \beta_j + v_{it} + \epsilon_{it}, \quad (39)$$

where

$$v_{it} = \sum_{l=1}^d \lambda_{il} f_{lt}. \quad (40)$$

Combining (39) with (40) and writing the model in matrix notation we get

$$Y_i = X_i \beta + F \Lambda_i^\top + \epsilon_i, \quad (41)$$

where  $Y_i = (y_{i1}, \dots, y_{iT})^\top$ ,  $X_i = (x_{i1}^\top, \dots, x_{iT}^\top)^\top$ ,  $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})^\top$ ,  $\Lambda_i = (\lambda_1, \dots, \lambda_n)^\top$  and  $F = (f_1, \dots, f_T)^\top$  with  $\lambda_i = (\lambda_{i1}, \dots, \lambda_{id})$ ,  $f_t = (f_{1t}, \dots, f_{dt})$ , and  $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})^\top$ .

The asymptotic properties of Bai's method rely, among others, on the following assumption:

$$\frac{1}{T} F^\top F \xrightarrow{p} \Sigma_F, \text{ as } T \rightarrow \infty, \quad (42)$$

where  $\Sigma_F$  is a fixed positive definite  $d \times d$  matrix. This allows for the factors to follow a deterministic time trend such as  $f_t = t/T$  or to be stationary dynamic processes such that  $f_t = \sum_{j=1}^{\infty} C_j e_{t-j}$ , where  $e_t$  are i.i.d. zero mean stochastic components. It is, however, important to note that such an assumption rules out a large class of non-stationary factors such as  $I(p)$  processes with  $p \geq 1$ .

#### 4.1. Model with known number of factors $d$

Bai (2009) proposes to estimate the model parameters  $\beta$ ,  $F$  and  $\Lambda_i$  by minimizing the following least squares objective function:

$$S(\beta, F, \Lambda_i) = \sum_{i=1}^n \|Y_i - X_i \beta - F \Lambda_i^\top\|^2. \quad (43)$$

For each given  $F$ , the OLS estimator of  $\beta$  can be obtained by

$$\hat{\beta}(F) = \left( \sum_{i=1}^n X_i^\top \mathcal{P}_d X_i \right)^{-1} \left( \sum_{i=1}^n X_i^\top \mathcal{P}_d Y_i \right), \quad (44)$$

where  $\mathcal{P}_d = I - F(F^\top F)^{-1} F^\top = I - FF^\top/T$ . If  $\beta$  is known,  $F$  can be estimated by using the first  $d$  eigenvectors  $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_d)$  corresponding to the first  $d$  eigenvalues of the empirical covariance matrix  $\hat{\Sigma} = (nT)^{-1} \sum_{i=1}^n w_i w_i^\top$ , where  $w_i = Y_i - X_i \beta$ . That is,

$$\hat{F}(\beta) = \sqrt{T} \hat{\gamma}.$$

The idea of Bai (2009) is to start with initial values for  $\beta$  or  $F$  and calculate the estimators iteratively. The method requires, however, the factor dimension  $d$  to be known, which is usually not the case in empirical applications.

A feasible estimator of (44) can be obtained by using an arbitrary large dimension  $d_{\max}$  greater than  $d$ . The factor dimension can be estimated subsequently by using the criteria of Bai and Ng (2002) to the remainder term  $Y_i = X_i \hat{\beta}(\hat{F}(d_{\max}))$ , as suggested by Bai (2009). This strategy can lead, however, to inefficient estimation and spurious interpretation of  $\beta$  due to over-parameterization.

#### 4.2. Model with unknown number of factors $d$

In order to estimate  $d$  jointly with  $\beta$ ,  $F$ , and  $\Lambda_i$ , [Bada and Kneip \(2014\)](#) propose to integrate a penalty term into the objective function to be globally optimized. In this case, the optimization criterion can be defined as a penalized least squares objective function of the form:

$$S(\beta, F, \Lambda_i, l) = \sum_{i=1}^n \|Y_i - X_i\beta - F\Lambda_i^\top\|^2 + lg_{nT} \quad (45)$$

The role of the additional term  $lg_{nT}$  is to pick up the dimension  $\hat{d}$ , of the unobserved factor structure. The penalty  $g_{nT}$  can be chosen according to [Bai and Ng \(2002\)](#). The estimation algorithm is based on the parameter cascading strategy of [Cao and Ramsay \(2010\)](#), which in this case can be described as follows:

1. Minimizing (45) with respect to  $\Lambda_i$  for each given  $\beta$ ,  $F$  and  $d$ , we get

$$\hat{\Lambda}_i^\top(\beta, F, d) = F^\top (Y_i - X_i\beta) / T. \quad (46)$$

2. Introducing (46) in (45) and minimizing with respect to  $F$  for each given  $\beta$  and  $d$ , we get

$$\hat{F}(\beta, d) = \sqrt{T}\hat{\gamma}(\beta, d), \quad (47)$$

where  $\hat{\gamma}(\beta, d)$  is a  $T \times d$  matrix that contains the first  $d$  eigenvectors corresponding to the first  $d$  eigenvalues  $\rho_1, \dots, \rho_d$  of the covariance matrix  $\hat{\Sigma} = (nT)^{-1} \sum_{i=1}^n w_i w_i^\top$  with  $w_i = Y_i - X_i\beta$ .

3. Reintegrating (47) and (46) in (45) and minimizing with respect to  $\beta$  for each given  $d$ , we get

$$\hat{\beta}(d) = \left( \sum_{i=1}^n X_i^\top X_i \right)^{-1} \left( \sum_{i=1}^n X_i^\top (Y_i - \hat{F}\hat{\Lambda}_i^\top(\hat{\beta}, d)) \right). \quad (48)$$

4. Optimizing (45) with respect to  $l$  given the results in (46), (47), and (48) allows us to select  $\hat{d}$  as

$$\hat{d} = \operatorname{argmin}_l \sum_{i=1}^n \|Y_i - X_i\hat{\beta} - \hat{F}\hat{\Lambda}_i^\top\|^2 + lg_{nT}, \quad \text{for all } l \in \{0, 1, \dots, d_{\max}\}.$$

The final estimators are obtained by alternating between an inner iteration to optimize  $\hat{\beta}(d)$ ,  $\hat{F}(d)$ , and  $\hat{\Lambda}_i(d)$  for each given  $d$  and an outer iteration to select the dimension  $\hat{d}$ . The updating process is repeated in its entirety till the convergence of all the parameters. This is why the estimators are called *entirely updated estimators* (Eup). In order to avoid over-estimation, [Bada and Kneip \(2014\)](#) propose to re-scale  $g_{nT}$  in each iteration stage with  $\hat{\sigma}^2 = \sum_i^n \|Y_i - X_i\hat{\beta} - \hat{F}\hat{\Lambda}_i^\top\|^2$  in stead of  $\hat{\sigma}^2(d_{\max})$ . Simulations show that such a calibration can improve the finite sample properties of the estimation method.

It is notable that the objective functions (43) and (45) are not globally convex. There is no guarantee that the iteration algorithm converges to the global optimum. Therefore, it is

important to choose reasonable starting values  $\hat{d}_{\text{start}}$  and  $\hat{\beta}_{\text{start}}$ . We propose to select a large dimension  $d_{\text{max}}$  and to start the iteration with the following estimate of  $\beta$ :

$$\hat{\beta}_{\text{start}} = \left( \sum_{i=1}^n X_i^\top (I - GG^\top) X_i \right)^{-1} \left( \sum_{i=1}^n X_i^\top (I - GG^\top) Y_i \right), \quad (49)$$

where  $G$  is the  $T \times d_{\text{max}}$  matrix of the eigenvectors corresponding to the first  $d_{\text{max}}$  eigenvalues of the augmented covariance matrix

$$\Gamma^{\text{Aug}} = \frac{1}{nT} \sum_{i=1}^n (Y_i, X_i)(Y_i^\top, X_i^\top)^\top.$$

The intuition behind these starting estimates relies on the fact that the unobserved factors cannot escape from the space spanned by the eigenvectors  $G$ . The projection of  $X_i$  on the orthogonal complement of  $G$  in (49) eliminates the effect of a possible correlation between the observed regressors and unobserved factors, which can heavily distort the value of  $\beta^0$  if it is neglected. [Greenaway-McGrevy, Han, and Sul \(2012\)](#) give conditions under which (49) is a consistent estimator of  $\beta$ . In order to avoid misspecifying the model through identifying factors that only exist in  $X_i$  and not  $Y_i$ , [Bada and Kneip \(2014\)](#) recommend to under-scale the starting common factors  $G_l$  that are highly correlated with  $X_i$ .

According to [Bai \(2009\)](#), the asymptotic distribution of the slope estimator  $\hat{\beta}(d)$  for known  $d$  is given by

$$\sqrt{nT}(\hat{\beta}(d) - \beta) \overset{a}{\sim} N(0, D_0^{-1} D_Z D_0^{-1}),$$

where  $D_0 = \text{plim}_{\frac{1}{nT}} \sum_{i=1}^n \sum_{t=1}^T Z_{it}^\top Z_{it}$  with  $Z_i = (Z_{i1}, \dots, Z_{iT})^\top = \mathcal{P}_d X_i - \frac{1}{n} \sum_{k=1}^n \mathcal{P}_d X_i a_{ik}$  and  $a_{ik} = \Lambda_i (\frac{1}{n} \sum_{i=1}^n \Lambda_i^\top \Lambda_i)^{-1} \Lambda_k^\top$ , and

- Case 1.  $D_Z = D_0^{-1} \sigma^2$  if the errors are i.i.d. with zero mean and variance  $\sigma^2$ ,
- Case 2.  $D_Z = \text{plim}_{\frac{1}{nT}} \sum_{i=1}^n \sigma_i^2 \sum_{t=1}^T Z_{it}^\top Z_{it}$ , where  $\sigma_i^2 = E(\epsilon_{it}^2)$  with  $E(\epsilon_{it}) = 0$ , if cross-section heteroskedasticity exists and  $n/T \rightarrow 0$ ,
- Case 3.  $D_Z = \text{plim}_{\frac{1}{nT}} \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} \sum_{t=1}^T Z_{it}^\top Z_{jt}$ , where  $\omega_{ij} = E(\epsilon_{it} \epsilon_{jt})$  with  $E(\epsilon_{it}) = 0$ , if cross-section correlation and heteroskedasticity exist and  $n/T \rightarrow 0$ ,
- Case 4.  $D_Z = \text{plim}_{\frac{1}{nT}} \sum_{t=1}^T \sigma_t^2 \sum_{i=1}^n Z_{it}^\top Z_{it}$ , where  $\sigma_t^2 = E(\epsilon_{it}^2)$  with  $E(\epsilon_{it}) = 0$ , if heteroskedasticity in the time dimension exists and  $T/n \rightarrow 0$ ,
- Case 5.  $D_Z = \text{plim}_{\frac{1}{nT}} \sum_{t=1}^T \sum_{s=1}^T \rho(t, s) \sum_{i=1}^n Z_{it}^\top Z_{is}$ , where  $\rho(t, s) = E(\epsilon_{it} \epsilon_{is})$  with  $E(\epsilon_{it}) = 0$ , if correlation and heteroskedasticity in the time dimension exist and  $T/n \rightarrow 0$ , and
- Case 6.  $D_Z = \text{plim}_{\frac{1}{nT}} \sum_{t=1}^T \sum_{i=1}^n \sigma_{it}^2 Z_{it}^\top Z_{is}$ , where  $\sigma_{it}^2 = E(\epsilon_{it}^2)$  with  $E(\epsilon_{it}) = 0$ , if heteroskedasticity in both time and cross-section dimensions exists with  $T/n^2 \rightarrow 0$  and  $n/T^2 \rightarrow 0$ .

In presence of correlation and heteroskedasticity in panels with proportional dimensions  $n$  and  $T$ , i.e.,  $n/T \rightarrow c > 0$ , the asymptotic distribution of  $\hat{\beta}(d)$  will be not centered at zero. This can lead to false inference when using the usual test statistics such as  $t$  and  $\chi^2$  statistic.

To overcome this problem, [Bai \(2009\)](#) proposes to estimate the asymptotic bias and correct the estimator as follows:

$$\hat{\beta}^*(d) = \hat{\beta}(d) - \frac{1}{n}\hat{B} - \frac{1}{T}\hat{C}, \quad (50)$$

where  $\hat{B}$  and  $\hat{C}$  are the estimators of

$$\begin{aligned} B &= -\left(\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T Z_{it}^\top Z_{it}\right)^{-1} \frac{1}{nT} \sum_{i=1}^n \sum_{k=1}^n (X_i - V_i)^\top F \left(\frac{1}{T} F^\top F\right)^{-1} W_{ik} \\ C &= -\left(\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T Z_{it}^\top Z_{it}\right)^{-1} \frac{1}{nT} \sum_{i=1}^n X_i^\top M_F \Omega F \left(\frac{1}{T} F^\top F\right)^{-1} \left(\frac{1}{n} \sum_{k=1}^n \Lambda_k^\top \Lambda_k\right)^{-1} \Lambda_i^\top \end{aligned}$$

respectively. Here,  $V_i = \frac{1}{n} \sum_{j=1}^n a_{ij} X_j$ ,  $W_{ik} = \left(\frac{1}{n} \sum_{j=1}^n \Lambda_j^\top \Lambda_j\right)^{-1} \Lambda_k^\top \frac{1}{T} \sum_{t=1}^T E(\epsilon_{it} \epsilon_{kt})$ , and  $\Omega = \frac{1}{n} \sum_{k=1}^n \Omega_k$  with

Case 7.  $\Omega_k$  is a  $T \times T$  diagonal matrix with elements  $\omega_{kt} = E(\epsilon_{kt}^2)$  if heteroskedasticity in both time and cross-section dimensions exist and  $n/T \rightarrow c > 0$  and,

Case 8.  $\Omega_k$  is a  $T \times T$  matrix with elements  $\Omega_{k,ts} = E(\epsilon_{kt} \epsilon_{ks})$  if correlation and heteroskedasticity in both time and cross-section dimensions exist and  $n/T \rightarrow c > 0$ .

In a similar context, [Bada and Kneip \(2014\)](#) prove that estimating  $d$  with the remaining model parameters does not affect the asymptotic properties of  $\hat{\beta}(d)$ . The asymptotic distribution of  $\hat{\beta} = \hat{\beta}(\hat{d})$  is given by

$$\sqrt{nT}(\hat{\beta} - \beta) \overset{a}{\sim} N(0, D_0^{-1} D_Z D_0^{-1})$$

under Cases 1–6, and

$$\sqrt{nT}(\hat{\beta}^* - \beta) \overset{a}{\sim} N(0, D_0^{-1} D_Z D_0^{-1})$$

under Cases 7–8, where  $\hat{\beta}^* = \hat{\beta}^*(\hat{d})$ .

The asymptotic variance of  $\hat{\beta}$  and the bias terms  $B$  and  $C$  can be estimated by replacing  $F$ ,  $\Lambda_i$ ,  $Z_{it}$ , and  $\epsilon_{it}$  with  $\hat{F}$ ,  $\hat{\Lambda}_i$ ,  $\hat{Z}_{it}$ , and  $\hat{\epsilon}_{it}$  respectively.

In presence of serial correlation (Cases 5 and 8), consistent estimators for  $D_Z$  and  $C$  can be obtained by using the usual heteroskedasticity and autocorrelation (HAC) robust limiting covariance. In presence of cross-section correlation (Case 3),  $D_Z$  is estimated by  $\hat{D}_Z = \frac{1}{mT} \sum_{i=1}^m \sum_{j=1}^m \sum_{t=1}^T \hat{Z}_{it}^\top \hat{Z}_{jt} \hat{\epsilon}_{it} \hat{\epsilon}_{jt}$ , where  $m = \sqrt{n}$ . If both cross-section and serial correlation exist (Case 8), we estimate the long-run covariance of  $\frac{1}{\sqrt{m}} \sum_{j=1}^m \hat{Z}_{it} \hat{\epsilon}_{it}$ .

### 4.3. Application

The above described methods are implemented in the function `Eup()`, which takes the following arguments:

```
Eup(formula, additive.effects = c("none", "individual", "time", "twoways"),
    dim.criterion = c("PC1", "PC2", "PC3", "BIC3", "IC1", "IC2", "IC3",
    "IPC1", "IPC2", "IPC3"), d.max = NULL, sig2.hat = NULL, factor.dim = NULL,
    double.iteration = TRUE, start.beta = NULL, max.iteration = 500,
    convergence = 1e-06, restrict.mode = c("restrict.factors",
    "restrict.loadings"), ...)
```

The arguments `additive.effects`, `d.max`, `sig2.hat`, and `restrict.mode` have the same roles as in `KSS()`; see Section 2.2. The argument `dim.criterion` specifies the dimensionality criterion to be used if `factor.dim` is left unspecified and defaults to `dim.criterion = "PC1"`. Setting the argument `double.iteration = FALSE` may speed up computations, because the updates of  $\hat{d}$  will be done simultaneously with  $\hat{F}$  without waiting for their inner convergences. However, in this case, the convergence of the parameters is less stable than in the default setting.

The argument `start.beta` allows us to give a vector of starting values for the slope parameters  $\beta_{\text{start}}$ . The maximal number of iteration and the convergence condition can be controlled by `max.iteration` and `convergence`.

In our application, we take first-order differences of the observed time series. This is because some factors show temporal trends, which can violate the stationarity condition (42); see Figure 2. We consider the following modified cigarettes model:

$$\Delta \ln(\text{Consumption}_{it}) = \beta_1 \Delta \ln(\text{Price}_{it}) + \beta_2 \Delta \ln(\text{Income}_{it}) + e_{it},$$

with

$$e_{it} = \sum_{l=1}^d \lambda_{il} f_{lt} + \epsilon_{it},$$

where  $\Delta x_t = x_t - x_{t-1}$ . In order to avoid notational mess, we use the same notation for the unobserved time-varying individual effects  $v_{it} = \sum_{l=1}^d \lambda_{il} f_{lt}$  as above in (20). The  $\Delta$ -transformation can be easily performed in R using the standard `diff()` function as follows:

```
R> d.l.Consumption <- diff(l.Consumption)
R> d.l.Price <- diff(l.Price)
R> d.l.Income <- diff(l.Income)
```

As previously mentioned for the `KSS()` function, the `formula` argument of the `Eup()` function takes balanced panel variables as  $T \times n$  dimensional matrices, where the number of rows has to be equal to the temporal dimension  $T$  and the number of columns has to be equal to the individual dimension  $n$ .

```
R> (Cigar.Eup <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
+   dim.criterion = "PC3"))
```

Call:

```
Eup.default(formula = d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
  dim.criterion = "PC3")
```

Coeff(s) of the Observed Regressor(s) :

```
  d.l.Price d.l.Income
-0.3140143  0.159392
```

Additive Effects Type: none

Dimension of the Unobserved Factors: 5

Number of iterations: 55

Inferences about the slope parameters can be obtained by using the `summary()` method. The type of correlation and heteroskedasticity in the idiosyncratic errors can be specified by choosing one of the corresponding Cases 1–8 described above using the argument `error.type = c(1, 2, 3, 4, 5, 6, 7, 8)`. In presence of serial correlations (Cases 5 and 8), the kernel weights required for estimating the long-run covariance can be externally specified by giving a vector of weights in the argument `kernel.weights`. By default, the function uses internally the linearly decreasing weights of Newey and West (1987) and a truncation at  $\lfloor \min\{\sqrt{n}, \sqrt{T}\} \rfloor$ . If Case 7 or 8 is chosen, the `summary()` method calculates the realization of the bias corrected estimators and gives appropriate inferences. The bias corrected coefficients can be called by using the `coef()` method on the object produced by `summary()`.

```
R> summary(Cigar.Eup)
```

Call:

```
Eup.default(formula = d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
  dim.criterion = "PC3")
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.147000	-0.013700	0.000889	0.014100	0.093300

Slope-Coefficients:

	Estimate	Std.Err	Z value	Pr(>z)
d.l.Price	-0.3140	0.0227	-13.90	< 2.2e-16 ***
d.l.Income	0.1590	0.0358	4.45	8.39e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Additive Effects Type: none

Dimension of the Unobserved Factors: 5

Residual standard error: 0.02804 on 957 degrees of freedom,  
R-squared: 0.7033

The summary output reports that "PC3" detects 5 common factors. The effect of the differenced log-real prices for cigarettes on the differenced log-sales is negative and amounts to  $-0.31$ . The estimated effect of the differenced real disposable log-income per capita is  $0.16$ . The estimated factors  $\hat{f}_{it}$  as well as the individual effects  $\hat{v}_{it}$  can be plotted using the `plot()` method for 'summary.Eup' objects. The corresponding graphics are shown in Figure 4.

```
R> plot(summary(Cigar.Eup))
```



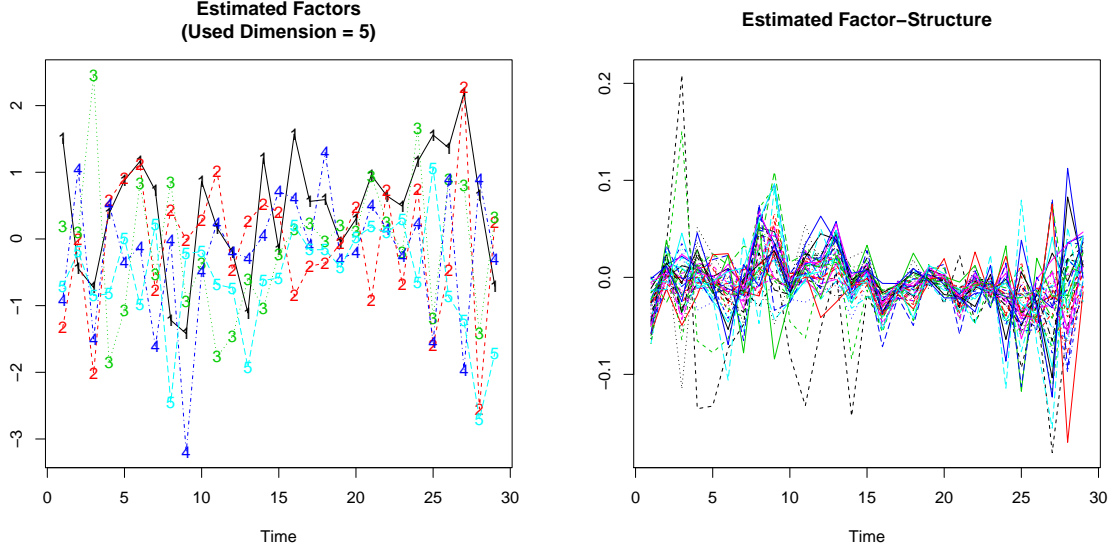


Figure 4: Left panel: Estimated factors  $\hat{f}_{1t}, \dots, \hat{f}_{7t}$ . Right panel: Estimated time-varying individual effects  $\hat{v}_{1t}, \dots, \hat{v}_{nt}$ .

## 5. Models with additive and interactive unobserved effects

Even though the classical additive "individual", "time", and "twoways" effects can be absorbed by the factor structure, there are good reasons to model them explicitly. On the one hand, if there are such effects in the true model, then neglecting them will result in non-efficient estimators; see [Bai \(2009\)](#). On the other hand, additive effects can be very useful for interpretation.

Consider now the following model:

$$y_{it} = \mu + \alpha_i + \theta_t + x_{it}^\top \beta + \nu_{it} + \epsilon_{it} \quad (51)$$

with

$$\nu_{it} = \begin{cases} v_{it} &= \sum_{l=1}^d \lambda_{il} f_{lt}, & \text{for the model of Bai (2009),} \\ v_i(t) &= \sum_{l=1}^d \lambda_{il} f_l(t), & \text{for the model of Kneip et al. (2012),} \end{cases}$$

where  $\alpha_i$  are time-constant individual effects and  $\theta_t$  is a common time-varying effect.

In order to ensure identification of the additional additive effects  $\alpha_i$  and  $\theta_t$ , we need the following further restrictions:

- (d)  $\sum_{i=1}^n \lambda_{il} = 0$  for all  $l \in \{1, \dots, d\}$ ,
- (e)  $\sum_{t=1}^T f_{lt} = 0$  for all  $l \in \{1, \dots, d\}$ ,
- (f)  $\sum_{i=1}^n \alpha_i = 0$ ,
- (g)  $\sum_{t=1}^T \theta_t = 0$ .

By using the classical within-transformations on the observed variables, we can eliminate the additive effects  $\alpha_i$  and  $\theta_t$ , such that

$$\dot{y}_{it} = \dot{x}_{it}^\top \beta + \nu_{it} + \epsilon_{it},$$

where  $\dot{y}_{it} = y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it} - \frac{1}{n} \sum_{i=1}^n y_{it} + \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n y_{it}$ ,  $\dot{x}_{it} = x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it} - \frac{1}{n} \sum_{i=1}^n x_{it} + \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n x_{it}$ , and  $\dot{\epsilon}_{it} = \epsilon_{it} - \frac{1}{T} \sum_{t=1}^T \epsilon_{it} - \frac{1}{n} \sum_{i=1}^n \epsilon_{it} + \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n \epsilon_{it}$ .

Note that restrictions (d) and (e) ensure that the transformation does not affect the time-varying individual effects  $\nu_{it}$ . The parameters  $\mu$ ,  $\alpha_i$  and  $\theta_t$  can be easily estimated in a second step once an estimate of  $\beta$  is obtained. Because of restrictions (d) and (e), the solution has the same form as the classical fixed effects model.

The parameters  $\beta$  and  $\nu_{it}$  can be estimated by the above introduced estimation procedures. All possible variants of model (51) are implemented in the functions `KSS()` and `Eup()`. The appropriate model can be specified by the argument `additive.effects = c("none", "individual", "time", "twoways")`:

"none"	$y_{it} = \mu + x_{it}^\top \beta + \nu_{it} + \epsilon_{it},$
"individual"	$y_{it} = \mu + \alpha_i + x_{it}^\top \beta + \nu_{it} + \epsilon_{it},$
"time"	$y_{it} = \mu + \theta_t + x_{it}^\top \beta + \nu_{it} + \epsilon_{it},$
"twoways"	$y_{it} = \mu + \alpha_i + \theta_t + x_{it}^\top \beta + \nu_{it} + \epsilon_{it}.$

The presence of  $\mu$  can be controlled by `-1` in the `formula` argument: a formula with `-1` refers to a model without intercept. However, for identification purposes, if a `twoways` model is specified, the presence `-1` in the `formula` will be ignored.

As an illustration, we continue with the application of the `KSS()` function in Section 2. The left panel of Figure 2 shows that the first common factor is nearly time-invariant. This motivates us to augment the model (20) with a time-constant additive effects  $\alpha_i$ . In this case, it is convenient to use an intercept  $\mu$ , which yields the following model:

$$\ln(\text{Consumption}_{it}) = \mu + \beta_1 \ln(\text{Price}_{it}) + \beta_2 \ln(\text{Income}_{it}) + \alpha_i + v_i(t) + \varepsilon_{it}, \quad (52)$$

where

$$v_i(t) = \sum_{l=1}^d \lambda_{il} f_l(t).$$

The estimation of the augmented model (52) can be done using the following lines of code.

```
R> Cigar2.KSS <- KSS(formula = 1.Consumption ~ 1.Price + 1.Income,
+   additive.effects = "individual")
R> (Cigar2.KSS.summary <- summary(Cigar2.KSS))
```

Call:

```
KSS.default(formula = 1.Consumption ~ 1.Price + 1.Income,
  additive.effects = "individual")
```

Residuals:

Min	1Q	Median	3Q	Max
-0.11	-0.01	0.00	0.01	0.12

Slope-Coefficients:

	Estimate	StdErr	z.value	Pr(>z)
(Intercept)	4.0500	0.1760	23.10	< 2.2e-16 ***
l.Price	-0.2600	0.0222	-11.70	< 2.2e-16 ***
l.Income	0.1570	0.0381	4.11	3.88e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Additive Effects Type: individual

Used Dimension of the Unobserved Factors: 5

Residual standard error: 0.000734 on 951 degrees of freedom

R-squared: 0.99

Again, the `plot()` method provides a useful visualization of the results.

```
R> plot(Cigar2.KSS.summary)
```

The "individual"-transformation of the data does not affect the estimation of the slope parameters, but reduces the estimated dimension from  $\hat{d} = 6$  to  $\hat{d} = 5$ . The remaining five common factors  $\hat{f}_1, \dots, \hat{f}_5$  correspond to those of model (20); see the middle panel of Figure 5. The estimated time-constant state-specific effects  $\alpha_i$  are shown in the left plot of Figure 5. The extraction of the  $\alpha_i$ 's from the factor structure yields a denser set of time-varying individual effects  $\hat{v}_i$  shown in the right panel of Figure 5.

### 5.1. Specification tests

Model specification is an important step for any empirical analysis. The **phtt** package is equipped with two types of specification tests: the first is a Hausman-type test appropriate for the model of Bai (2009). The second one examines the existence of a factor structure in Bai's model as well as in the model of Kneip *et al.* (2012).

#### *Testing the sufficiency of classical additive effects*

For the case in which the estimated number of factors amounts to one or two ( $1 \leq \hat{d} \leq 2$ ), it is interesting to check whether or not these factors can be interpreted as classical "individual", "time", or "twoways" effects. Bai (2009) considers the following testing problem:

$$\begin{aligned} H_0: & \quad v_{it} = \alpha_i + \theta_t \\ H_1: & \quad v_{it} = \sum_{l=1}^2 \lambda_{il} f_{lt} \end{aligned}$$

The model with factor structure, as described in Section 4, is consistent under both hypotheses. However, it is less efficient under  $H_0$  than the classical within estimator, while the latter is

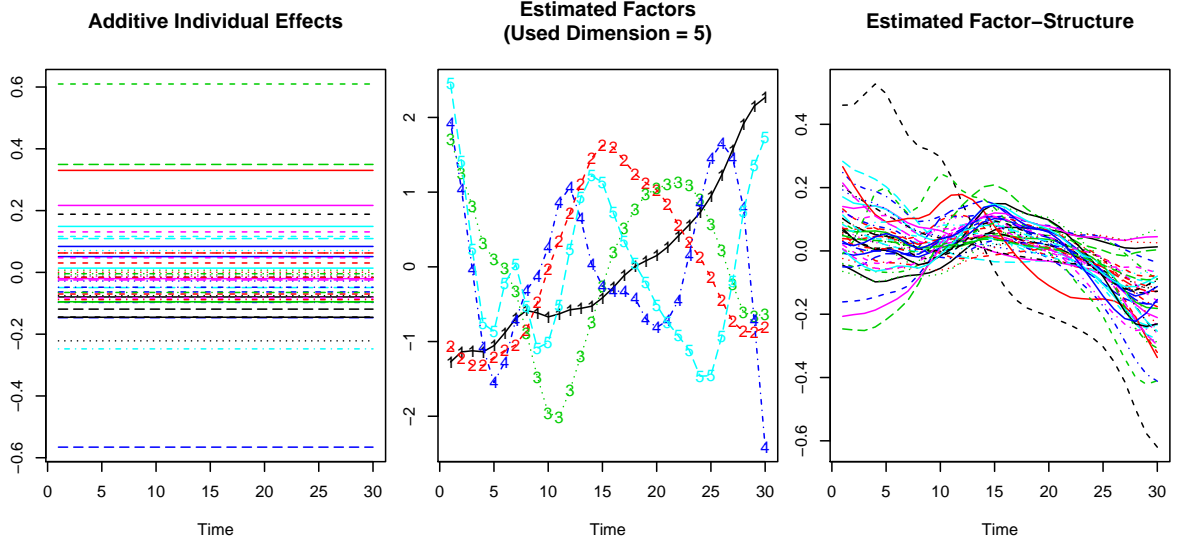


Figure 5: Left panel: Estimated time-constant state-specific effects  $\hat{\alpha}_1, \dots, \hat{\alpha}_n$ . Middle panel: Estimated common factors  $\hat{f}_1(t), \dots, \hat{f}_5(t)$ . Right panel: Estimated time-varying individual effects  $\hat{v}_1(t), \dots, \hat{v}_n(t)$ .

inconsistent under  $H_1$  if  $x_{it}$  and  $v_{it}$  are correlated. These conditions are favorable for applying the Hausman test:

$$J_{\text{Bai}} = nT \left( \hat{\beta} - \hat{\beta}_{\text{within}} \right) \Delta^{-1} \left( \hat{\beta} - \hat{\beta}_{\text{within}} \right) \stackrel{a}{\sim} \chi_P^2, \quad (53)$$

where  $\hat{\beta}_{\text{within}}$  is the classical within least squares estimator,  $\Delta$  is the asymptotic variance of  $\sqrt{nT} \left( \hat{\beta} - \hat{\beta}_{\text{within}} \right)$ ,  $P$  is the vector-dimension of  $\beta$ , and  $\chi_P^2$  is the  $\chi^2$ -distribution with  $P$  degrees of freedom.

The null hypothesis  $H_0$  can be rejected, if  $J_{\text{Bai}} > \chi_{P,1-\alpha}^2$ , where  $\chi_{P,1-\alpha}^2$  is the  $(1-\alpha)$ -quantile of the  $\chi^2$  distribution with  $P$  degrees of freedom.

Under i.i.d. errors,  $J_{\text{Bai}}$  can be calculated by replacing  $\Delta$  with its consistent estimator

$$\hat{\Delta} = \left( \left( \frac{1}{nT} \sum_{i=1}^n Z_i^\top Z_i \right)^{-1} - \left( \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T x_{it} x_{it}^\top \right)^{-1} \right) \hat{\sigma}^2, \quad (54)$$

where

$$\hat{\sigma}^2 = \frac{1}{nT - (n+T)\hat{d} - P + 1} \sum_{i=1}^n \sum_{t=1}^T (y_{it} - x_{it}^\top \hat{\beta} - \sum_{l=1}^{\hat{d}} \hat{\lambda}_{il} \hat{f}_{lt})^2. \quad (55)$$

The residual variance estimator  $\hat{\sigma}^2$  is chosen here, since it is supposed to be consistent under the null as well as the alternative hypothesis. The idea behind this trick is to avoid negative definiteness of  $\hat{\Delta}$ . But notice that even with this construction, the possibility of getting a negative definite variance estimator cannot be excluded. As an illustration, consider the case in which the true number of factors is greater than the number of factors used under

the alternative hypothesis, i.e., the true  $d > 2$ . In such a case, the favorable conditions for applying the test can be violated, since the iterated least squares estimator  $\hat{\beta}$  is computed with  $\hat{d} \leq 2$  and can be inconsistent under both hypotheses. To avoid such a scenario, we recommended to the user to calculate  $\hat{\beta}$  with a large dimension  $d_{\max}$  instead of  $\hat{d} \leq 2$ .

The test is implemented in the function `checkSpecif()`, which takes the following arguments:

```
R> checkSpecif(obj1, obj2, level = 0.05)
```

The argument `level` is used to specify the significance level. The arguments `obj1` and `obj2` take both objects of class `Eup` produced by the function `Eup()`:

`obj1` Takes an ‘Eup’ object from an estimation with "individual", "time", or "twoways" effects and a factor dimension equal to  $d = 0$ ; specified as `factor.dim = 0`.

`obj2` Takes an ‘Eup’ object from an estimation with "none"-effects and a large factor dimension  $d_{\max}$ ; specified with the argument `factor.dim`.

If the test statistic is negative (due to the negative definiteness of  $\hat{\Delta}$ ), the `checkSpecif()` prints an error message.

```
R> twoways.obj <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
+   factor.dim = 0, additive.effects = "twoways")
R> not.twoways.obj <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
+   factor.dim = 2, additive.effects = "none")
R> checkSpecif(obj1 = twoways.obj, obj2 = not.twoways.obj, level = 0.01)
```

```
Error in checkSpecif(obj1 = twoways.obj, obj2 = not.twoways.obj,
level = 0.01):
The assumptions of the test are not fulfilled.
The (unobserved) true number of factors is probably greater than 2.
```

Notice that the Hausman test of [Bai \(2009\)](#) assumes the within estimator to be inconsistent under the alternative hypothesis, which requires  $x_{it}$  to be correlated with  $v_{it}$ . If this assumption is violated, the test can suffer from power to reject the null hypothesis, since the within estimator becomes consistent under both hypotheses.

[Bai \(2009\)](#) discusses in his supplementary material another way to check whether a classical panel data with fixed additive effects is sufficient to describe the data. His idea consists of estimating the factor dimension after eliminating the additive effects as described in Section 5. If the obtained estimate of  $d$  is zero, the additive model can be considered as a reasonable alternative for the model with factor structure. But note that this procedure cannot be considered as a formal testing procedure, since information about the significance level of the decision are not provided.

An alternative test for the sufficiency of a classical additive effects model can be given by manipulating the test proposed by [Kneip et al. \(2012\)](#) as described in the following section.

*Testing the existence of common factors*

This section is concerned with testing the existence of common factors. In contrast to the Hausman-type statistic discussed above, the goal of this test is not merely to decide which model specification is more appropriate for the data, but rather to test in general the existence of common factors beyond the possible presence of additional classical "individual", "time", or "twoways" effects in the model.

This test relies on using the dimensionality criterion proposed by [Kneip et al. \(2012\)](#) to test the following hypothesis after eliminating eventual additive "individual", "time", or "twoways" effects:

$$\begin{aligned} H_0: & \quad d = 0 \\ H_1: & \quad d > 0 \end{aligned}$$

Under  $H_0$  the slope parameters  $\beta$  can be estimated by the classical within estimation method. In this simple case, the dimensionality test of [Kneip et al. \(2012\)](#) can be reduced to the following test statistic:

$$J_{\text{KSS}} = \frac{n \operatorname{tr}(\hat{\Sigma}_w) - (n-1)(T-1)\hat{\sigma}^2}{\sqrt{2n(T-1)\hat{\sigma}^2}} \stackrel{a}{\sim} N(0, 1),$$

where  $\hat{\Sigma}_w$  is the covariance matrix of the within residuals. The reason for this simplification is that under  $H_0$  there is no need for smoothing, which allows us to set  $\kappa = 0$ .

We reject  $H_0: d = 0$  at a significance level  $\alpha$ , if  $J_{\text{KSS}} > z_{1-\alpha}$ , where  $z_{1-\alpha}$  is the  $(1-\alpha)$ -quantile of the standard normal distribution. It is important to note that the performance of the test depends heavily on the accuracy of the variance estimator  $\hat{\sigma}^2$ . We propose to use the variance estimators (15) or (55), which are consistent under both hypotheses as long as  $\hat{d}$  is greater than the unknown dimension  $d$ . Internally, the test procedure sets  $\hat{d} = \mathbf{d.max}$  and  $\sigma^2$  as in (55).

This test can be performed for 'Eup' as well as for 'KSS' objects using the function `checkSpecif()` leaving the second argument `obj2` unspecified. In the following, we apply the test for both models.

For the model of [Bai \(2009\)](#):

```
R> Eup.obj <- Eup(d.l.Consumption ~ -1 + d.l.Price + d.l.Income,
+   additive.effects = "twoways")
R> checkSpecif(Eup.obj, level = 0.01)
```

```
-----
Testing the Presence of Interactive Effects
Test of Kneip, Sickles, and Song (2012)
-----
```

H0: The factor dimension is equal to 0.

Test-Statistic	p-value	crit.-value	sig.-level
13.29	0.00	2.33	0.01

For the model of [Kneip et al. \(2012\)](#):

```
R> KSS.obj <- KSS(l.Consumption ~ -1 + l.Price + l.Income,
+   additive.effects = "twoways")
R> checkSpecif(KSS.obj, level = 0.01)
```

```
-----
Testing the Presence of Interactive Effects
Test of Kneip, Sickles, and Song (2012)
-----
```

H0: The factor dimension is equal to 0.

Test-Statistic	p-value	crit.-value	sig.-level
104229.55	0.00	2.33	0.01

The null hypothesis  $H_0: d = 0$  can be rejected for both models at a significance level  $\alpha = 0.01$ .

## 6. Interpretation

This section is intended to outline an exemplary interpretation of the panel model (52), which is estimated by the function `KSS()` in Section 5. The interpretation of models estimated by the function `Eup()` can be done accordingly. For convenience sake, we re-write the model (52) in the following:

$$\ln(\text{Consumption}_{it}) = \mu + \beta_1 \ln(\text{Price}_{it}) + \beta_2 \ln(\text{Income}_{it}) + \alpha_i + v_i(t) + \varepsilon_{it},$$

where

$$v_i(t) = \sum_{l=1}^d \lambda_{il} f_l(t).$$

A researcher, who chooses the panel models proposed by Kneip *et al.* (2012) or Bai (2009), will probably find them attractive due to their ability to control for very general forms of unobserved heterogeneity. Beyond this, a further great advantage of these models is that the time-varying individual effects  $v_i(t)$  provide a valuable source of information about the *differences* between the individuals  $i$ . These differences are often of particular interest as, e.g., in the literature on stochastic frontier analysis.

The left panel of Figure 5 shows that the different states  $i$  have considerable different time-constant levels  $\hat{\alpha}_i$  of cigarette consumption. A classical further econometric analysis could be to regress the additive individual effects  $\hat{\alpha}_i$  on other time-constant variables, such as the general populations compositions, the cigarette taxes, etc.

The right panel of Figure 5 shows the five estimated common factors  $\hat{f}_1(t), \dots, \hat{f}_5(t)$ . It is good practice to start the interpretation of the single common factors with an overview about their importance in describing the differences between the  $v_i(t)$ 's, which is reflected in the variances of the individual loadings parameters  $\hat{\lambda}_{il}$ . A convenient depiction is the quantity of variance-shares of the individual loadings parameters on the total variance of the loadings parameters

$$\text{coef}(\text{Cigar2.KSS})\$Var.shares.of.loadings.param[l] = V(\hat{\lambda}_{il}) / \sum_{k=1}^{\hat{d}} V(\hat{\lambda}_{ik}),$$



Common factor	Share of total variance of $v_i(t)$
$\hat{f}_1(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[1]</code> = 66.32%
$\hat{f}_2(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[2]</code> = 24.28%
$\hat{f}_3(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[3]</code> = 5.98%
$\hat{f}_4(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[4]</code> = 1.92%
$\hat{f}_5(t)$	<code>coef(Cigar2.KSS)\$Var.shares.of.loadings.param[5]</code> = 1.50%

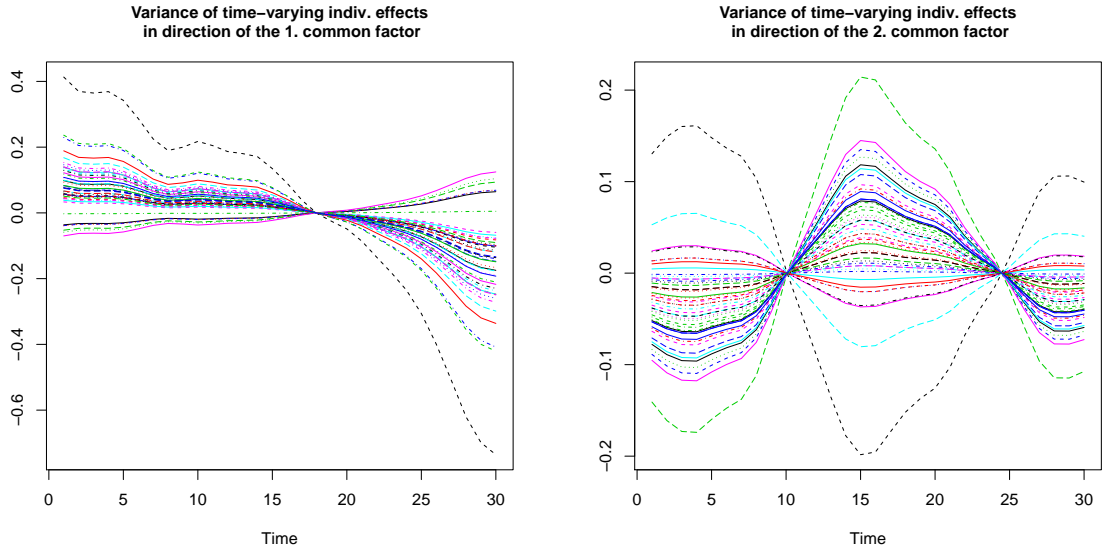
Table 1: List of the variance shares of the common factors  $\hat{f}_1(t), \dots, \hat{f}_5(t)$ .

Figure 6: Left panel: Visualization of the differences of the time-varying individual effects  $v_i(t)$  in the direction of the first factor  $\hat{f}_1(t)$  (i.e.,  $\hat{\lambda}_{i1}\hat{f}_1(t)$ ). Right panel: Visualization of the differences of the time-varying individual effects  $v_i(t)$  in the direction of the second factor  $\hat{f}_2(t)$  (i.e.,  $\hat{\lambda}_{i2}\hat{f}_2(t)$ ).

which is shown for all common functions  $\hat{f}_1(t), \dots, \hat{f}_5(t)$  in Table 1.

The values in Table 1 suggest to focus on the first two common factors, which explain together about 90% of the total variance of the time-varying individual effects  $\hat{v}_i(t)$ .

The first two common factors

$$\begin{aligned} \text{coef(Cigar2.KSS)}\$Common.factors[,1] &= \hat{f}_1(t) \quad \text{and} \\ \text{coef(Cigar2.KSS)}\$Common.factors[,2] &= \hat{f}_2(t) \end{aligned}$$

are plotted as black and red lines in the middle panel of Figure 5. Figure 6 visualizes the differences of the time-varying individual effects  $v_i(t)$  in the direction of the first common factor (i.e.,  $\hat{\lambda}_{i1}\hat{f}_1(t)$ ) and in the direction of the second common factor (i.e.,  $\hat{\lambda}_{i2}\hat{f}_2(t)$ ). As for the time-constant individual effects  $\hat{\alpha}_i$  a further econometric analysis could be to regress the individual loadings parameters  $\hat{\lambda}_{i1}$  and  $\hat{\lambda}_{i2}$  on other explanatory time-constant variables.

Generally, for both models proposed by Kneip *et al.* (2012) and Bai (2009) the time-varying

individual effects

$$\nu_{it} = \sum_{l=1}^d \lambda_{il} f_{lt}$$

can be interpreted as it is usually done in the literature on factor models. An important topic that is not covered in this section is the rotation of the common factors. Often, the common factors  $f_l$  can be interpreted economically only after the application of an appropriate rotation scheme for the set of factors  $\hat{f}_1, \dots, \hat{f}_{\hat{d}}$ . The latter can be done, e.g., using the function `varimax()` from the **stats** package. Alternatively, many other rotation schemes can be found in the **GPArotation** package (R Core Team 2014; Bernaards and Jennrich 2005). Sometimes, it is also preferable to standardize the individual loadings parameters instead of the common factors as it is done, e.g., in Ahn, Lee, and Schmidt (2001). This can be done by choosing `restrict.mode = "restrict.loadings"` in the functions `KSS()` and `Eup()` respectively.

## 7. Summary

This paper introduces the R package **phtt** for the new class of panel models proposed by Bai (2009) and Kneip *et al.* (2012). The two main functions of the package are the `Eup()` function for the estimation procedure proposed in Bai (2009) and the `KSS()` function for the estimation procedure proposed in Kneip *et al.* (2012). Both of the main functions are supported by the usual `print()`, `summary()`, `plot()`, `coef()` and `residuals()` methods. While parts of the method of Bai (2009) are available for commercially available software packages, the estimation procedure proposed by Kneip *et al.* (2012) is not available elsewhere. A further remarkable feature of our **phtt** package is the `OptDim()` function, which provides an easy access to many different dimensionality criteria proposed in the literature on factor models. The usage of the functions is demonstrated by a real data application.

## Acknowledgments

The authors wish to thank the referees for their many helpful comments and suggestions that greatly improved the paper. The work of Dominik Liebl was supported by the IAP Research NetworkP7/06 of the Belgian State (Belgian Science Policy).

## References

- Ahn SC, Horenstein AR (2013). “Eigenvalue Ratio Test for the Number of Factors.” *Econometrica*, **81**(3), 1203–1227.
- Ahn SC, Lee YH, Schmidt P (2001). “GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects.” *Journal of Econometrics*, **101**(2), 219–255.
- Ahn SC, Lee YH, Schmidt P (2013). “Panel Data Models with Multiple Time-Varying Individual Effects.” *Journal of Econometrics*, **174**(1), 1–14.
- Alessi L, Barigozzi M, Capasso M (2010). “Improved Penalization for Determining the Number of Factors in Approximate Factor Models.” *Statistics & Probability Letters*, **80**(23–24), 1806–1813.

- Bada O, Kneip A (2014). “Parameter Cascading for Panel Models with Unknown Number of Unobserved Factors: An Application to the Credit Spread Puzzle.” *Computational Statistics & Data Analysis*, **76**, 95–115.
- Bada O, Liebl D (2014). *phtt: Panel Data Analysis with Heterogeneous Time Trends*. R package version 3.1.2, URL <http://CRAN.R-project.org/package=phtt>.
- Bai J (2004). “Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data.” *Journal of Econometrics*, **122**(1), 137–183.
- Bai J (2009). “Panel Data Models with Interactive Fixed Effects.” *Econometrica*, **77**(4), 1229–1279.
- Bai J, Kao C, Ng S (2009). “Panel Cointegration with Global Stochastic Trends.” *Journal of Econometrics*, **149**(1), 82–99.
- Bai J, Ng S (2002). “Determining the Number of Factors in Approximate Factor Models.” *Econometrica*, **70**(1), 191–221.
- Baltagi BH (2005). *Econometric Analysis of Panel Data*. 3rd edition. John Wiley & Sons.
- Baltagi BH, Levin D (1986). “Estimating Dynamic Demand for Cigarettes Using Panel Data: The Effects of Bootlegging, Taxation and Advertising Reconsidered.” *The Review of Economics and Statistics*, **68**(1), 148–155.
- Baltagi BH, Li D (2004). “Prediction in the Panel Data Model with Spatial Correlation.” In *Advances in Spatial Econometrics: Methodology, Tools and Applications*, 1st edition, pp. 283–295. Springer-Verlag.
- Bates D, Maechler M, Bolker B, Walker S (2014). *lme4: Linear Mixed-Effects Models Using Eigen and S4*. R package version 1.1-7, URL <http://CRAN.R-project.org/package=lme4>.
- Bernaards CA, Jennrich RI (2005). “Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis.” *Educational and Psychological Measurement*, **65**(5), 676–696.
- Boor CD (2001). *A Practical Guide to Splines*. Applied Mathematical Series, revised edition. Springer-Verlag.
- Cao J, Ramsay JO (2010). “Linear Mixed-Effects Modeling by Parameter Cascading.” *Journal of the American Statistical Association*, **105**(489), 365–374.
- Craven P, Wahba G (1978). “Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation.” *Numerische Mathematik*, **31**(4), 377–403.
- Croissant Y, Milla G (2008). “Panel Data Econometrics in R: The **plm** Package.” *Journal of Statistical Software*, **27**(2), 1–43. URL <http://www.jstatsoft.org/v27/i02/>.
- Greenaway-McGrevy R, Han C, Sul D (2012). “Asymptotic Distribution of Factor Augmented Estimators for Panel Regression.” *Journal of Econometrics*, **169**(1), 48–53.

- Hallin M, Liška R (2007). “Determining the Number of Factors in the General Dynamic Factor Model.” *Journal of the American Statistical Association*, **102**(478), 603–617.
- Kneip A, Sickles RC, Song W (2012). “A New Panel Data Treatment for Heterogeneity in Time Trends.” *Econometric Theory*, **28**(3), 590–628.
- Millo G, Piras G (2012). “**splm**: Spatial Panel Data Models in R.” *Journal of Statistical Software*, **47**(1), 1–38. URL <http://www.jstatsoft.org/v47/i01/>.
- Newey WK, West KD (1987). “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, **55**(3), 703–708.
- Onatski A (2010). “Determining the Number of Factors from Empirical Distribution of Eigenvalues.” *The Review of Economics and Statistics*, **92**(4), 1004–1016.
- Pesaran HM (2006). “Estimation and Inference in Large Heterogeneous Panels with a Multi-factor Error Structure.” *Econometrica*, **74**(4), 967–1012.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2014). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-117, URL <http://CRAN.R-project.org/package=nlme>.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- The MathWorks Inc (2012). *MATLAB – The Language of Technical Computing, Version 7.14*. Natick, Massachusetts. URL <http://www.mathworks.com/products/matlab/>.

**Affiliation:**

Oualid Bada  
Statistische Abteilung  
University of Bonn  
Adenauerallee 24-26  
53113 Bonn, Germany  
E-mail: [bada@uni-bonn.de](mailto:bada@uni-bonn.de)

Dominik Liebl  
Université Libre de Bruxelles  
CP 210, local O.9.213  
Bd. du Triomphe  
1050 Bruxelles, Belgique  
E-mail: [dominik.liebl@ulb.ac.be](mailto:dominik.liebl@ulb.ac.be)